SYNOPSIS

Search for New Physics in Proton-Proton Collisions Using the CMS Detector and Simulation-Based Studies for High Energy Physics Experiments

SUBMITTED BY

Malay Neerav

(Class Roll No: PAT-2204)

(PhD Course Work Exam Roll No: PHY-02/22)

UNIVERSITY DEPARTMENT OF PHYSICS

Bhupendra Narayan Mandal University

November, 2025

1. Title

Search for New Physics in Proton-Proton Collisions Using the CMS Detector and Simulation-Based Studies for High Energy Physics Experiments

2. Introduction

The Standard Model (SM) of particle physics has been highly successful in describing elementary particles and their interactions. However, several unexplained phenomena—including the existence of dark matter, the origin of neutrino masses, the hierarchy problem, and matter–antimatter asymmetry—indicate that the SM is incomplete. These open questions motivate the search for new physics beyond the Standard Model (BSM).

The Large Hadron Collider (LHC) at CERN provides the highest-energy proton-proton collisions currently available, creating opportunities to probe BSM physics at the energy frontier. The Compact Muon Solenoid (CMS) detector, one of the two large general-purpose detectors at the LHC, plays a central role in exploring rare processes, studying the Higgs boson, and searching for signals of new physics.

The proposed research aims to conduct a CMS Run-3 data analysis focused on signatures such as mono-Higgs or mono-Z final states, which are sensitive to a wide range of BSM scenarios including dark matter models, Two-Higgs Doublet Models (2HDM), and Higgs-portal extensions. In addition to data analysis, the study will involve Geant4-based detector simulation research to understand how detector geometry affects performance, and how advanced machine learning techniques can enhance particle identification and reconstruction.

3. Review of Literature

The search for physics beyond the Standard Model (BSM) has been a central objective of the Large Hadron Collider (LHC) research programme. A vast body of literature highlights the theoretical motivations, experimental strategies, and detector-level challenges associated with such searches. This section provides an overview of several key areas relevant to the proposed study: mono-Higgs and mono-Z signatures, dark matter and Higgs-portal models, extended Higgs sectors, detector simulation using Geant4, and machine-learning-based improvements in particle reconstruction.

Mono-X Searches at the LHC: Mono-X signatures, where a Higgs boson or Z boson recoils against invisible particles, are well-motivated channels for dark matter searches at colliders. Early theoretical work discussed the feasibility of mono-Higgs production as a portal to dark matter, particularly through gluon–gluon fusion or associated production mechanisms [1]. Experimental studies by the CMS and ATLAS collaborations have placed significant limits on such scenarios using $\sqrt{s} = 13$ TeV data, exploring final states such as $H \to \gamma\gamma$, $H \to b\bar{b}$, and $Z \to \ell^+\ell^-$ [2, 3]. The mono-Z channel, particularly in the leptonic decay modes, provides a clean final state with strong sensitivity to axial-vector mediator models [4].

Dark Matter and Higgs-Portal Models: Dark matter models motivated by astrophysical evidence often predict weakly interacting massive particles (WIMPs) that couple to the Standard Model through mediator particles. Higgs-portal models are among the simplest and most theoretically appealing scenarios, where a scalar singlet interacts with the SM via the Higgs field [5]. Collider searches for Higgs-portal dark matter typically rely on missing transverse momentum (p_T^{miss}) signatures accompanied by Higgs production. Similarly, simplified models with vector or axial vector mediators have been used to interpret mono-X results in a model-independent manner [6]. These frameworks have

guided much of the current LHC dark matter search strategies.

Extended Higgs Sectors and 2HDM: The Two-Higgs Doublet Model (2HDM) and its variants, including 2HDM+S and 2HDM+a, introduce additional scalar or pseudoscalar fields that can couple to dark matter or modify Higgs boson production rates. These models predict rich phenomenology, including resonant Higgs pair production, modified couplings, and new invisible decay channels [7]. CMS analyses involving mono-Higgs signatures have been interpreted within these frameworks to constrain parameter spaces relevant to dark matter interactions [8]. Theoretical reviews emphasise that mono-Higgs and mono-Z channels can probe mediator masses from the electroweak scale to several TeV [9].

Detector Simulation and CMS Performance Studies: Accurate simulation of particle interactions in detector materials is crucial for interpreting LHC data. Geant4 remains the standard toolkit for modelling electromagnetic and hadronic processes, optical photon propagation, and detector geometry effects [10]. Numerous CMS publications highlight the dependence of detector performance—such as energy resolution, timing resolution, and photon identification efficiency—on detector segmentation, materials, and sensor configurations [11]. Simulation-driven optimization plays a key role in guiding detector upgrades.

Machine Learning in High Energy Physics: Machine learning (ML) methods have transformed particle reconstruction and classification in collider experiments. Deep neural networks, boosted decision trees, and graph neural networks have demonstrated superior performance over traditional rule-based algorithms in tasks such as jet tagging, photon–π^0 separation, and pile-up mitigation [12]. CMS has implemented ML-based photon identification and jet substructure tools, achieving significant gains in background rejection [13]. Studies also show that ML can improve calorimeter energy regression and track

reconstruction, making it highly relevant for both physics analysis and simulation studies [14].

The reviewed literature establishes a strong foundation for the proposed research. Mono-Higgs and mono-Z channels remain promising avenues for discovering new physics with CMS Run 3 data. Theoretical models such as Higgs-portal dark matter and extended Higgs sectors provide well-motivated frameworks for interpreting search results. Detector simulation studies using Geant4 continue to play a crucial role in performance optimization, while machine learning has emerged as a powerful tool for enhancing reconstruction and analysis sensitivity. Together, these elements form a coherent basis for a comprehensive Ph.D. study involving physics analysis, detector simulation, and advanced reconstruction techniques.

4. Statement of the Problem

Despite the tremendous success of the Standard Model (SM) in describing known particles and interactions, several key observations—including the existence of dark matter, the origin of neutrino masses, and the matter–antimatter asymmetry—remain unexplained. These shortcomings strongly motivate the search for physics beyond the Standard Model (BSM). Experimental searches at the LHC have so far not yielded conclusive evidence of new physics, but many theoretically motivated scenarios remain only partially explored, particularly in final states involving a Higgs or Z boson in association with large missing transverse momentum.

The CMS experiment, with its excellent particle reconstruction capabilities and the high-luminosity data from Run 3, offers new opportunities to probe such signatures. However, significant challenges persist. First, the sensitivity to BSM signals depends strongly on accurate background modelling and optimized reconstruction techniques. Secondly, many theoretical models—including Higgs-portal dark matter, simplified mediator frameworks, and ex-

tended Higgs sectors such as 2HDM—predict subtle deviations that require refined analysis strategies. Third, the performance of detector components, such as calorimeters and scintillator-based systems, can influence the ability to reconstruct photons, jets, and missing energy with high precision. Understanding how different detector geometries or configurations impact these observables remains an open problem.

Additionally, traditional reconstruction techniques face limitations in separating closely related signatures, such as distinguishing prompt photons from background sources like π^0 decays. Recent advances in machine learning offer promising solutions, but their full potential has not yet been systematically explored within the context of BSM searches or detector performance studies.

Therefore, there is a clear need for a comprehensive investigation that combines: (i) a CMS Run 3 data analysis targeting mono-Higgs or mono-Z final states; (ii) theoretical interpretations within multiple BSM frameworks; (iii) Geant4-based simulation studies on detector geometry and performance; and (iv) machine-learning-driven improvements in particle identification and reconstruction.

Addressing these interconnected challenges will contribute to enhancing the sensitivity of new physics searches and to improving detector design and analysis methodologies for future high-energy physics experiments.

5. Objectives

General Objective

The primary objective of this research is to search for signatures of physics beyond the Standard Model (BSM) using proton–proton collision data from the CMS experiment during Run 3, with emphasis on Higgs or Z-associated production channels and simulation-based studies for detector performance optimization.

Specific Objectives

The study aims to achieve the following specific objectives:

1. To perform a detailed physics analysis of CMS Run 3 data focusing on mono-Higgs and mono-Z final states, which are sensitive to BSM scenarios involving dark matter or extended Higgs sectors.

2. To interpret the results within multiple theoretical frameworks, including:

   - Higgs-portal dark matter models,
   - simplified mediator models (vector, axial-vector, or scalar mediators),
   - Two-Higgs-Doublet Models (2HDM) and their extensions.

3. To develop a robust event-selection strategy and optimize background estimation using both Monte Carlo simulations and data-driven techniques for accurate extraction of potential BSM signals.

4. To conduct Geant4-based detector simulation studies aimed at understanding how variations in detector geometry, materials, and segmentation influence spatial resolution, light yield, and particle reconstruction efficiency.

5. To explore and implement machine-learning (ML) methods for improving particle identification and reconstruction, particularly in tasks such as:

   - photon–π^0 discrimination,
   - jet or object-level classification,
   - regression-based energy or position reconstruction.

6. Methodology

The methodology of the proposed research integrates CMS data analysis, detector-simulation studies, and machine-learning-based reconstruction techniques to conduct a comprehensive search for new physics in proton–proton collisions.

The approach is designed to systematically investigate mono-Higgs or mono-Z final states, interpret results within multiple BSM frameworks, and explore detector-performance improvements relevant to future high-energy physics experiments.

6.1 CMS Data Analysis Approach

The analysis will utilize proton–proton collision data recorded by the CMS detector during Run 3. Events will be reconstructed using the Particle Flow (PF) algorithm, which provides optimal identification of photons, leptons, jets, and missing transverse momentum. Event selection will be performed using appropriate triggers and object-identification criteria recommended by the CMS collaboration.

Monte Carlo simulations for both Standard Model backgrounds and BSM signal hypotheses will be generated using `MadGraph5_aMC@NLO` and `Pythia8`, followed by full CMS detector simulation. Background processes such as $Z(\nu\nu)$+jets, W+jets, top-quark production, and QCD multijet events will be estimated using a combination of MC predictions and data-driven control samples. Statistical analysis will be carried out using the CMS `Combine` tool, employing likelihood fits to derive signal strengths and exclusion limits.

6.2 Interpretation in New Physics Models

To provide a complete understanding of the results, the selected final states will be interpreted in several theoretically motivated BSM frameworks. These include Higgs-portal dark matter models, simplified mediator models, and extended Higgs-sector theories such as 2HDM and 2HDM+a. Parameter scans and benchmark models will be used to evaluate the sensitivity of the analysis to a broad range of mediator masses and couplings.

6.3 Detector Simulation Using Geant4

A complementary component of the study involves detector-performance investigations using Geant4. Multiple scintillator-tile geometries (square, hexagonal, trapezoidal, and monolithic designs) will be simulated to evaluate optical-photon transport, light yield, timing response, and spatial resolution. Various SiPM configurations will also be modelled to study their influence on signal collection and reconstruction. These studies aim to quantify how detector design choices affect performance metrics critical for calorimetry, photon detection, and precision measurements.

6.4 Machine Learning for Reconstruction and Identification

Machine-learning techniques will be incorporated to enhance particle identification and event reconstruction. Convolutional neural networks, boosted decision trees, and graph neural networks will be developed for tasks such as photon–π^0 separation, background suppression, jet substructure analysis, and calorimeter energy regression. Performance will be evaluated using ROC curves, efficiency–purity metrics, and significance improvement. The best-performing models will be validated using both simulated samples and appropriate control regions in CMS data.

6.5 Synthesis and Outcome

Results from CMS data analysis, theoretical model interpretation, detector simulations, and ML-based reconstruction tools will be synthesised to enhance the overall sensitivity to possible new physics signatures. The findings will also contribute to improved detector-design strategies and advanced reconstruction methodologies relevant for future high-energy physics experiments.

7. Significance of the Study

This study holds significant importance in advancing searches for physics beyond the Standard Model by exploiting CMS Run 3 data in mono-Higgs and mono-Z channels, which remain highly promising but underexplored. The integration of Geant4-based detector performance studies provides valuable insights for future detector design and optimization. Additionally, the development of machine-learning-based reconstruction techniques enhances particle identification and background rejection, improving overall analysis sensitivity. The combined outcomes will contribute to more precise measurements, stronger constraints on new physics models, and methodological advancements beneficial for upcoming high-energy physics experiments.

References

[1] F. D'Eramo et al., *JHEP*, 2016.

[2] CMS Collaboration, "Search for dark matter produced in association with a Higgs boson," *Phys. Rev. D*, 2018.

[3] ATLAS Collaboration, "Mono-Higgs searches at $\sqrt{s} = 13$ TeV," *JHEP*, 2019.

[4] CMS Collaboration, "Search for new physics in mono-Z final states," *JHEP*, 2021.

[5] Arcadi et al., "Higgs-portal dark matter models," *Eur. Phys. J. C*, 2020.

[6] Abercrombie et al., "Dark Matter Benchmark Models for LHC," *Phys. Dark Universe*, 2015.

[7] Branco et al., "Theory and phenomenology of 2HDM," *Phys. Rept.*, 2012.

[8] CMS Collaboration, "Constraints on 2HDM+a models from mono-Higgs," CMS PAS HIG-20-003.

[9] P. Tunney et al., *JHEP*, 2018.

[10] S. Agostinelli et al., "Geant4—a simulation toolkit," *Nucl. Instrum. Meth. A*, 2003.

[11] CMS Collaboration, "Performance of CMS calorimeters," *JINST*, 2017.

[12] Guest et al., "Deep Learning in HEP," *Ann. Rev. Nucl. Part. Sci.*, 2018.

[13] CMS Collaboration, "ML-based photon identification," CMS DP-2020/046.

[14] Paganini et al., "Calorimeter shower simulation with GANs," *Phys. Rev. Lett.*, 2018.

Proposed Chapterization

This Ph.D. thesis will be organized into the following chapters:

Chapter 1: Introduction

This chapter will introduce the motivation for searching for physics beyond the Standard Model and highlight the limitations of current theoretical frameworks. It will describe the role of the Large Hadron Collider (LHC) and the CMS detector in probing high-energy interactions, with emphasis on mono-Higgs and mono-Z signatures as promising channels for new physics searches. The research problem, objectives, scope, and structure of the thesis will also be presented.

Chapter 2: Literature Review

This chapter will provide an extensive review of earlier studies on mono-X searches, Higgs-portal and mediator-based dark matter models, extended Higgs sectors, and related CMS and ATLAS analyses. It will also summarize key works on detector simulation using Geant4 and the use of machine learning in particle identification and event reconstruction. The chapter will identify research gaps and motivate the present investigation.

Chapter 3: Methodology

This chapter will describe the methodological framework of the research, including CMS Run 3 data analysis procedures, Monte Carlo generation of Standard Model backgrounds and BSM signals, event reconstruction using the Particle Flow algorithm, and background estimation techniques. It will also detail statistical analysis using the CMS Combine tool, interpretation strategies within various BSM models, detector-performance studies using Geant4, and the development of machine-learning models for reconstruction and photon–π^0 separation. Tools, datasets, and evaluation metrics will be clearly outlined.

Chapter 4: Results and Discussion

This chapter will present the results of the CMS data analysis, including event yields, kinematic distributions, background estimates, and statistical limits on new physics models. It will also include findings from Geant4-based detector-performance studies and machine-learning-driven reconstruction improvements. Comparative evaluations will be provided, and all results will be interpreted in the context of existing experimental and theoretical literature.

Chapter 5: Conclusion and Future Scope

This chapter will summarize the key findings of the thesis and highlight contributions to new physics searches, detector optimization, and data-analysis methodologies. Broader implications for high-energy physics experiments will be discussed. The chapter will also present potential directions for future research, including enhanced analyses with higher-luminosity data, further detector-geometry optimization, and advanced ML-based reconstruction techniques.

References

This section will list all cited works, following the appropriate citation format required for the thesis.